

ARTICLE

Deep Learning–Based Detection of Ancient Agricultural Terraces Using Multisensor Data Fusion: A Case Study from the Bozburun Peninsula, Turkey

Emin Atabey Peker 

Middle East Technical University, Graduate School of Social Sciences Settlement Archaeology, Üniversiteler Mahallesi, Ankara, Turkey

 This research article was awarded an Open Materials badge for transparent practices. See the Data Availability Statement for details.

Email: eminatabeypeker@gmail.com

(Received 22 August 2025; revised 21 November 2025; accepted 23 November 2025; first published online 30 March 2026)

Abstract

The manual identification of ancient agricultural terraces is time-consuming and subjective, limiting large-scale archaeological landscape documentation. This study applies deep learning to detect ancient terraces in the Bozburun Peninsula, southwestern Turkey, a historically significant Hellenistic landscape. Four U-Net–based architectures were implemented—early, intermediate, and late fusion, along with an RGB-only baseline—integrating high-resolution aerial imagery (30 cm) and digital elevation models (DEMs) across 193 km². Sixteen manually digitized areas (37.8 ha) produced 256 training patches (512 × 512 px). The early fusion model that combined spectral and topographic data achieved the best performance (IoU = 0.754; accuracy = 85.9%). Monte Carlo evaluation confirmed its robustness. Spatial analysis showed that 89.8% of detected terraces lie below 300 m elevation, mainly on 10°–20° slopes with north-northwest orientation, in agreement with previous archaeological observations. Compared with expert digitization, the model yielded higher precision (87.4% vs. 79.3%), while experts achieved higher recall (94.3% vs. 76.6%). Applied to the full peninsula, the model mapped 2,517 ha of terraces. Validation using an existing archaeological dataset (Demirciler 2014) enabled direct comparison between automated and expert-based interpretations. The results indicate the potential of deep learning for terrace detection in Mediterranean landscapes and outline a methodological framework for documenting threatened cultural heritage.

Özet

Manuel olarak antik tarım teraslarını belirlemek zaman alıcı ve öznel, bu da büyük ölçekli arkeolojik peyzaj dokümantasyonunu kısıtlamaktadır. Bu çalışma, tarihsel olarak önemli bir Helenistik peyzaj olan güneybatı Türkiye'deki Bozburun Yarımadası'ndaki antik terasları tespit etmek için derin öğrenmeyi uygulamaktadır. Yüksek çözünürlüklü hava fotoğrafı (30 cm) ve sayısal yükseklik modellerini 193 km² alan genelinde entegre eden dört U-Net tabanlı mimari (erken, ara ve geç füzyon ile sadece RGB içeren bir temel çizgi) uygulandı. Manuel olarak sayısallaştırılmış on altı alan (37.8 ha), 256 eğitim yaması (512 × 512 piksel) üretti. Spektral ve topografik verileri birleştiren erken füzyon modeli en iyi performansı elde etti (IoU = 0.754; doğruluk = 85.9%). Monte Carlo değerlendirmesi, modelin sağlamlığını doğruladı. Önceki arkeolojik gözlemlerle uyumlu olacak şekilde, uzamsal analiz tespit edilen terasların %89.8'inin 300 metrenin altında bir yükseklikte, esas olarak kuzey-kuzeybatı yönelimli 10°–20° eğimlerde bulunduğunu gösterdi. Uzman sayısallaştırmasıyla karşılaştırıldığında, model daha yüksek kesinlik (Precision) sağladı (%87.4'e karşılık %79.3), buna karşın uzmanlar daha yüksek geri çağırma (Recall) elde etti (%94.3'e karşılık %76.6). Tüm yarımada'ya uygulandığında, model 2,517 ha teras alanı haritaladı. Mevcut bir arkeolojik veri kümesi kullanılarak yapılan doğrulama (Demirciler

2014), otomatik ve uzman tabanlı yorumlar arasında doğrudan karşılaştırma yapılmasını sağladı. Sonuçlar, derin öğrenmenin Akdeniz peyzajlarındaki teras tespiti için potansiyelini göstermekte ve tehdit altındaki kültürel mirasın belgelenmesi için metodolojik bir çerçeve sunmaktadır.

Keywords: agricultural terraces; Bozburun Peninsula; deep learning; digital archaeology; remote sensing

Anahtar kelimeler: tarım terasları; Bozburun Yarımadası; derin öğrenme; dijital arkeoloji; uzaktan algılama

Agricultural terraces are enduring landscape modifications that allowed ancient societies to cultivate steep slopes, reduce erosion, and retain water in arid and semiarid environments (Krahtopoulou and Frederick 2008). Composed of treads and risers, they represent a key form of ancient agro-ecological adaptation. Found across the Andes, East Asia, and the Mediterranean, terraces have served not only agricultural but also environmental, cultural, and social purposes (Guttmann-Bond 2019).

The Bozburun Peninsula in southwestern Anatolia illustrates the strategic use of terracing in rugged terrain. During the Hellenistic period, it became part of the Rhodian Peraia—a network of mainland territories administered by Rhodes from the third century BCE onward. This administrative and economic integration encouraged intensified agriculture, particularly through terrace construction and the production of wine and olive oil (Demirciler et al. 2022). Despite its peripheral and erosion-prone topography, Bozburun developed into a productive agricultural landscape connected to regional and Mediterranean trade routes. Archaeological evidence from settlements such as ancient Tymnos reflects this integration: amphora stamps and inscriptions point to continuous economic activity within Rhodian administrative structures (Demirciler 2017).

From a landscape-archaeological perspective, Bozburun's terraces provide essential evidence for understanding the agricultural economy and spatial organization of Hellenistic Caria. Settlement clusters, pressing installations, water management features, and fortified acropoleis illustrate a well-organized territorial model adapted to local environmental constraints. The combination of fortresses, sacred land leasing, and diversified crop production—including cereals, grapes, olives, figs, and almonds—indicates a resilient agro-economic system that persisted into Late Antiquity (Demirciler et al. 2023; Kuban 2008). This context makes Bozburun an important case for examining rural transformation under Rhodian rule and the long-term sustainability of terraced landscapes.

Despite their archaeological importance and broad distribution, the systematic documentation of ancient agricultural terraces faces significant methodological challenges. Traditional archaeological surveys, although providing valuable insights into construction techniques and historical settings, are time-consuming and labor-intensive when applied to extensive terrace systems.

To overcome these limitations, various remote sensing approaches have been developed. Sofia and colleagues (2014) introduced the Slope Local Length of Auto-Correlation (SLLAC) method to distinguish terraced landscapes from natural formations, showing that terraces exhibit more regular patterns than natural slopes. Capolupo and colleagues (2018) advanced terrace mapping through Object-Based Image Analysis (OBIA), producing binary maps of terraced and nonterraced areas from aerial imagery. Cucchiari and colleagues (2020) combined Terrestrial Laser Scanning (TLS) and Structure from Motion (SfM) to monitor terrace morphology under complex topographic conditions.

Agapiou and colleagues (2019) introduced the concept of *virtual constellations*, emphasizing how the coordinated use of optical and radar satellite sensors can enhance archaeological landscape analysis by integrating complementary spectral and structural information. These semi-automated techniques improved efficiency and demonstrated the potential of computational tools in archaeological mapping, yet they continued to rely on extensive human interpretation, limiting their scalability for landscape-scale analysis. Preservation also varies widely across terrace systems. Many have been altered by erosion, abandonment, or modern land use, posing additional detection challenges that conventional and semi-automated methods address inconsistently. Before the introduction of deep learning, Santos and colleagues (2019) used a Support Vector Machine (SVM) with Local Binary Pattern (LBP) descriptors to classify steep-slope vineyards in Portugal's Douro region, demonstrating that texture-based features can effectively delineate vineyard structures even without large training datasets.

Tubog and colleagues (2025) applied an SVM-based model combining lidar-derived terrain data and high-resolution orthophotos for terrace detection in the Roya Valley, France, using an early fusion strategy that achieved high mapping accuracy. Gravel-Miguel and colleagues (2025) later implemented an early-fusion U-Net framework integrating multiple lidar-derived visualizations (Simple Local Relief Model, slope, Terrain Ruggedness Index, and Positive Openness) for terrace segmentation in Georgia, USA, achieving high recall even with limited training data. Together, these studies illustrate the gradual transition from classical to machine learning approaches, culminating in recent applications of deep learning and data fusion for large-scale archaeological landscape analysis.

Recent developments in deep learning have improved the automated detection of agricultural terraces from remote sensing data, addressing many of the scalability and objectivity limitations of earlier methods. Do and colleagues (2019) provided one of the first neural network applications for terrace extraction using 5 m RapidEye imagery in Vietnam. Their comparison of pixel-based and object-based approaches with an eight-layer feed-forward network showed that pixel-based classification performed better owing to reduced information loss from averaging.

Expanding to multisensor approaches, Zhao and colleagues (2021) implemented a U-Net architecture combining remote sensing imagery with digital elevation models (DEMs) from China's Loess Plateau. Postprocessing constraints derived from slope and terrain features (5°–25° positive slopes) reduced noise and separated terraces from surrounding farmlands. Figueiredo and colleagues (2022) applied a U-Net with an Inception ResNet V2 encoder to detect curved terrace vineyards in Portugal's Alto Douro wine region, demonstrating the effectiveness of transfer learning for terrace mapping in complex topographies. Building on this, Figueiredo and colleagues (2023) extended the method to identify vine rows within terraces using UAV RGB imagery and DEM data, addressing the added challenge of steep and curved terrain.

Wang and colleagues (2022) introduced a semantic model fusion framework that combined aerial imagery and lidar-based terrain data. Their late fusion approach merged the probability outputs of separately trained U-Net and DeepLabv3+ models ($\alpha = 0.5$), balancing precision and recall to achieve the highest ranking in the International AI Archaeology Challenge. Lu and colleagues (2023) proposed an early fusion strategy through their Deep Learning-based Terrace Extraction Model (DLTEM), using a U-Net++ architecture that integrated RGB imagery with eight DEM-derived features as multichannel input. The model demonstrated high accuracy at 1.89 m resolution in the Loess Plateau, substantially improving on coarser-resolution products.

Ciglić and colleagues (2024) demonstrated terrace recognition based solely on lidar-derived DEMs using a U-Net architecture, showing effective performance in densely forested areas of Slovenia, where optical data are often unreliable. Zhao and colleagues (2024) developed NLDF-Net, a feature-level fusion model that incorporates attention mechanisms through Asymmetric Nonlocal and Dual Fusion Blocks. The framework achieved strong results and represents a refined stage in deep learning-based terrace mapping. Table 1 summarizes the principal studies from 2019 to 2024, outlining their datasets, model architectures, and fusion strategies.

Although recent advances have improved automated terrace detection, an important question remains: how do different data fusion strategies perform in archaeological settings, particularly in Mediterranean landscapes? Data fusion involves combining information from multiple sources to enhance detection accuracy. In remote sensing, this integration can occur at different stages—by merging inputs at the start (early fusion), combining features during processing (intermediate fusion), or integrating model outputs at the end (late fusion; Qiu et al. 2022). Early fusion combines all input data (e.g., RGB and topography) into a single multichannel image before learning begins. Intermediate fusion processes different data types through separate encoders and merges them during feature extraction. Late fusion keeps data streams separate and combines their predictions only at the final decision stage.

This study investigates that question by developing and evaluating U-Net-based deep learning models for identifying ancient agricultural terraces in the Bozburun Peninsula using high-resolution remote sensing data. It compares early, intermediate, and late fusion strategies that integrate spectral (RGB) and topographic (DEM, slope, aspect) variables, along with an RGB-only baseline to assess the added

Table 1. Summary of Previous Terrace Detection Studies Highlighting Their Data Sources, Analytical Methods, Fusion Strategies, and Key Contributions.

Study	Data Source	Method	Fusion Type	Key Contribution
Sofia and colleagues 2014	DEM	SLLAC morphometric analysis	None	Quantified terrace regularity using terrain autocorrelation.
Capolupo and colleagues 2018	Aerial photos	OBIA	None	Automated terrace mapping from aerial imagery.
Cucchiaro and colleagues 2020	TLS + SfM	3D data fusion	Early	Monitored terrace morphodynamics via multisensor data.
Agapiou and colleagues 2019	Sentinel-1/2 + DEM	OBIA + GIS metrics	Early	Linked radar-optical data to detect terraces.
Santos and colleagues 2019	Satellite imagery	SVM + LBP	None	Texture-based classification of vineyard terraces.
Do and colleagues 2019	RapidEye imagery	Feed-forward DNN	None	Early neural network for terrace extraction.
Zhao and colleagues 2021	Imagery + DEM	U-Net + slope filter	Early	Combined spectral-topographic data for terrace mapping.
Figueiredo and colleagues 2022	Google Earth imagery	U-Net (Inception ResNet V2)	None	Transfer learning for curved terrace vineyards.
Ciglič and colleagues 2024	LiDAR DEM	U-Net segmentation	None	DEM-only deep learning terrace recognition.
Zhao and colleagues 2024	Imagery + DEM	NLDF-Net	Intermediate	Attention-based fusion for high-accuracy detection.
Tubog and colleagues 2025	LiDAR DTM + orthophotos	SVM	Early	Lidar-optical fusion for terrace mapping.
Gravel-Miguel and colleagues 2025	LiDAR visualizations	U-Net segmentation	Early	Large-scale terrace detection for heritage sites.
Figueiredo and colleagues 2023	UAV RGB + DEM	Hybrid ML/DL	Early	UAV data fusion for vine-row identification.
Lu and colleagues 2023	Satellite + DEM + vegetation	U-Net++ (DLTEM)	Early	Multisource fusion terrace map at 1.89 m resolution.
Wang and colleagues 2022	Aerial + LiDAR	U-Net + DeepLabv3+	Late	Late-fusion ensemble for ancient terrace detection.

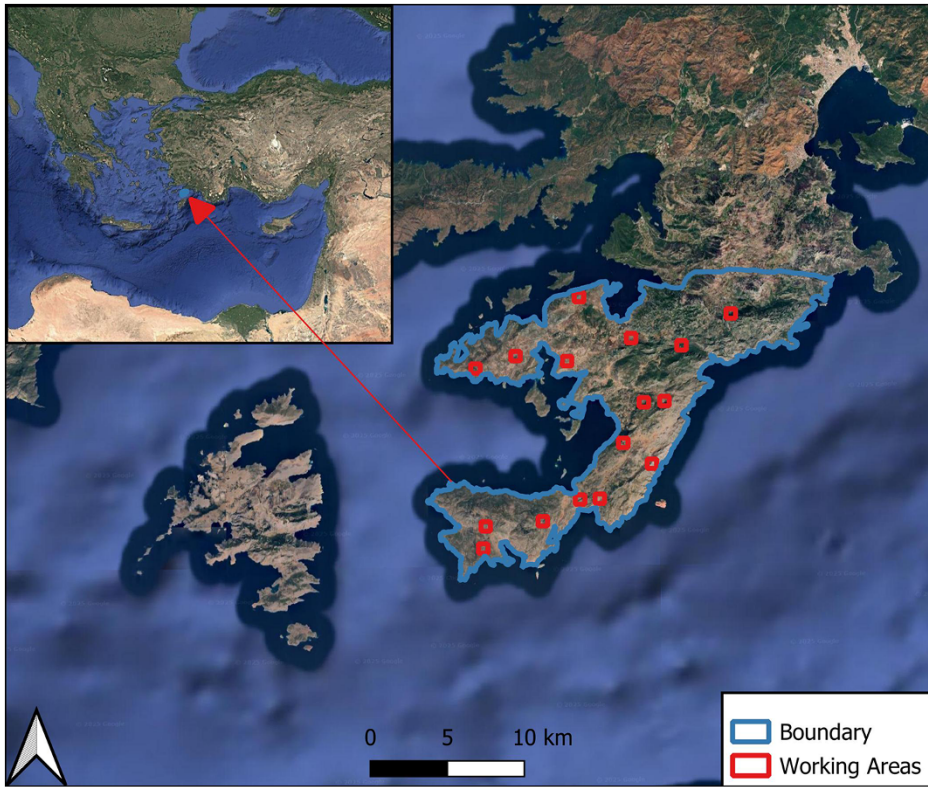


Figure 1. Spatial distribution of 16 working units (37.8 ha total) selected for training data across the Bozburun Peninsula.

value of topographic information. Model performance is evaluated using multiple metrics to ensure a balanced assessment.

The availability of a detailed, expert-digitized terrace dataset for the region (Demirciler 2014) provides a valuable opportunity for thorough validation—something rarely possible in archaeological machine learning research. The results contribute to improving automated terrace mapping and demonstrate how deep learning can support large-scale cultural landscape documentation.

Materials and Methods

Data and Preprocessing

This study used high-resolution aerial imagery (30 cm, RGB, acquired in 2022) and DEMs (30 cm, acquired in 2022) covering the 193 km² Bozburun Peninsula in southwestern Turkey. The peninsula contains extensive terrace systems dating mainly to the Hellenistic period. Slope and aspect derivatives were generated from the DEM using QGIS 3.34 to characterize topographic conditions relevant to terrace identification. This approach focused on terrain morphology rather than vegetation patterns, ensuring that terrace detection relied primarily on geomorphological characteristics.

Sixteen representative sample areas (37.8 ha in total) were selected across the study area to ensure balanced representation of terraced and nonterraced landscapes (Figure 1). These areas were digitized to create ground-truth labels, taking into account terrace morphology and preservation state. The resulting vector data were converted to raster format to produce binary classification labels (terrace vs. nonterrace) for model training.

To visualize the datasets used in terrace detection, Figure 2 presents the RGB orthomosaic together with the DEM, slope, and aspect layers. These layers provide the key spectral and topographic inputs used in subsequent fusion experiments.

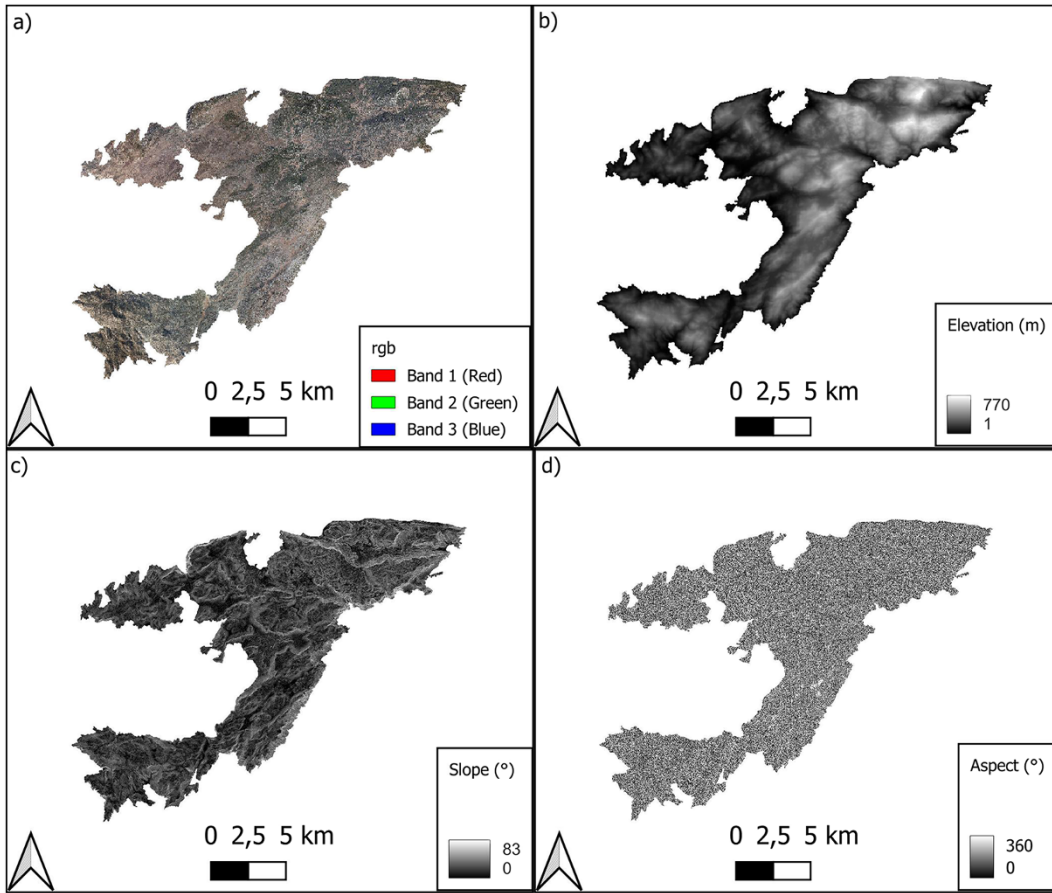


Figure 2. Core input datasets of the Bozburun Peninsula: (a) RGB orthomosaic; (b) DEM (m); (c) slope (°); (d) aspect (°). All layers share identical spatial extent with standardized scale and north orientation.

All datasets (RGB, DEM, slope, aspect, and labels) were divided into 512×512 px tiles, generating 256 image patches for deep learning experiments (Table 2). Each raster layer was exported as NumPy array files (.npy) for efficient processing and model compatibility.

To ensure consistent scaling across input channels and to address the circular nature of aspect values ($0^\circ = 360^\circ$), all inputs were normalized as follows. RGB values were scaled to the (0, 1) range. Elevation and slope were normalized by dividing each pixel value by the maximum elevation (781 m) and maximum slope (84°) within the study area (Equation 1). Aspect values were transformed into circular coordinates to avoid angular discontinuity, using sine and cosine components (Equation 2), since aspect represents directional orientation on a 0° – 360° circle, where 0° and 360° correspond to the same direction. This circular encoding preserves directional continuity and prevents artificial breaks that would occur if values were linearly normalized between 0 and 1.

$$s' = s/s_{max} \quad (1)$$

$$a'_{sin} = (\sin\theta + 1)/2, a'_{cos} = (\cos\theta + 1)/2 \quad (2)$$

Table 2. Specifications of 512 × 512 Image Patches used for Model Training.

Data Type	Dimensions	Description
RGB	512,512,3	Three-channel color image
DEM	512,512,1	Digital Elevation Model
Slope	512,512,1	Terrain slope information
Aspect	512,512,1	Terrain aspect information
Label	512,512,1	Manually digitized

Dataset splitting followed a Monte Carlo cross-validation approach (Xu and Liang 2001) with 10 random initializations to ensure robust evaluation. Monte Carlo validation assesses model generalization by repeatedly training on randomly sampled subsets of data and averaging performance across runs. In practical terms, this approach reduces overfitting by testing the model on different subsets of data and averaging the results. For each split, 63% of the data (161 patches) was used for training, 27% (69 patches) for validation, and 10% (26 patches) for independent testing. The same random seeds were applied across all fusion strategies to ensure comparability between models (Table 3). These repeated random splits provided a statistically reliable measure of model stability and performance variation across different fusion configurations.

Deep Learning Architecture

To evaluate the contribution of topographic data to terrace detection and to compare different data integration strategies, four U-Net-based architectures were implemented. An RGB-only baseline was first established to assess detection performance using spectral information alone. Three additional fusion approaches were then designed to integrate spectral (RGB) and topographic (DEM, slope, aspect) data: early fusion (input-level integration), intermediate fusion (feature-level integration), and late fusion (decision-level integration).

Different imaging modalities provide complementary information: RGB imagery captures color and texture characteristics, while elevation data contribute topographic context that enhances detection accuracy through multimodal integration (Boulahia et al. 2021; Qiu et al. 2022). This comparative framework allows quantitative evaluation of both the added value of topographic data and the relative performance of different fusion strategies for archaeological terrace detection. Similar systematic fusion analyses have also been effective in medical imaging and remote sensing contexts (Raju et al. 2025).

U-Net-Based Architecture

The U-Net architecture (Ronneberger et al. 2015) served as the foundation for all experiments owing to its proven efficiency in image segmentation tasks and suitability for limited archaeological datasets. The characteristic U-shaped design consists of a contracting encoder path that captures contextual information at multiple scales and an expanding decoder path that reconstructs detailed spatial predictions. Skip connections between encoder and decoder layers preserve fine spatial details while maintaining semantic consistency, which is particularly useful for identifying subtle terrace boundaries in heterogeneous landscapes.

RGB-Only Baseline

The RGB-only model functions as a reference to evaluate the effectiveness of spectral data alone. Each 512 × 512 input image (three-channel RGB) was processed through a standard U-Net with normalized values in the (0–1) range. The encoder comprises four convolutional blocks (64, 128, 256, and 512 filters), each containing two convolutional layers followed by batch normalization, ReLU (Rectified Linear Unit) activation, and dropout (rate = 0.4). A bridge layer with 1,024 filters connects the encoder and decoder. The decoder mirrors the encoder with four upsampling stages and skip connections to recover spatial details. The final layer produces binary segmentation maps using softmax activation for terrace vs. background classification.

Table 3. Dataset Split Distribution for Model Training and Evaluation.

Split	Number of Patches	Percentage (%)
Training	161	63
Validation	69	27
Test	26	10
Total	256	100

Early Fusion

The early fusion model integrates spectral and topographic data at the input stage through a unified seven-channel tensor combining RGB (three channels) with four normalized topographic derivatives (elevation, slope, aspect_sin, aspect_cos). All modalities are processed simultaneously from the first convolutional layer, allowing interaction between spectral and terrain features during early feature extraction. The encoder follows the same configuration as the baseline (64–512 channels). This approach is designed to enhance learning of complementary patterns between RGB textures and topographic gradients, aiding the detection of terrace structures that may not be clearly visible in a single data source.

Intermediate Fusion

Two parallel encoders process RGB (three channels) and topographic data (four channels: elevation, slope, aspect_sin, aspect_cos) separately through four convolutional blocks (64–512 filters). At the bridge, encoder outputs are concatenated (1,024 channels) and processed by a 1,024-filter block. The decoder uses four upsampling stages where, at each level, skip connections from both encoders are fused together before integration with upsampled features, enabling hierarchical multimodal fusion.

Late Fusion

The late fusion approach maintains fully independent U-Net architectures for spectral and topographic data, combining their outputs only at the decision level. One network processes the RGB imagery, while the other processes the topographic channels. Each produces binary segmentation maps through softmax activation. The two prediction maps are concatenated and passed through a 1×1 convolutional layer with softmax activation to produce the final terrace map. This design maintains independence between modalities while allowing adaptive weighting of their outputs based on local image context.

The four fusion architectures are illustrated in Figure 3, which demonstrates the distinct data integration strategies employed in this study. Figure 3a shows the RGB-only baseline processing spectral information through a standard U-Net, while Figure 3b depicts the early fusion approach combining all input channels at the network input. Figure 3c illustrates the intermediate fusion strategy with parallel encoders and feature-level integration, and Figure 3d presents the late fusion architecture with independent networks and decision-level combination. These architectural variations enable systematic evaluation of how different fusion strategies affect terrace detection performance in archaeological contexts.

Network Components

All fusion architectures used the same U-Net building blocks so that any differences in results reflect only how the data are combined. We used the ReLU activation function after each convolution and batch-normalization step. ReLU sets negative values to zero and keeps positive values unchanged, which makes training faster and helps the network learn sharp terrace boundaries (Equation 3).

$$f(x) = \max(0, x) \quad (3)$$

To reduce overfitting, we combined dropout and batch normalization. Dropout was applied in the encoder (rate 0.4 in the first convolution of each block and 0.2 in the second), and batch normalization was used before every ReLU to stabilize training. All models were trained with Jaccard (Intersection over Union [IoU]) loss, which directly optimizes the overlap between predicted and reference terrace masks

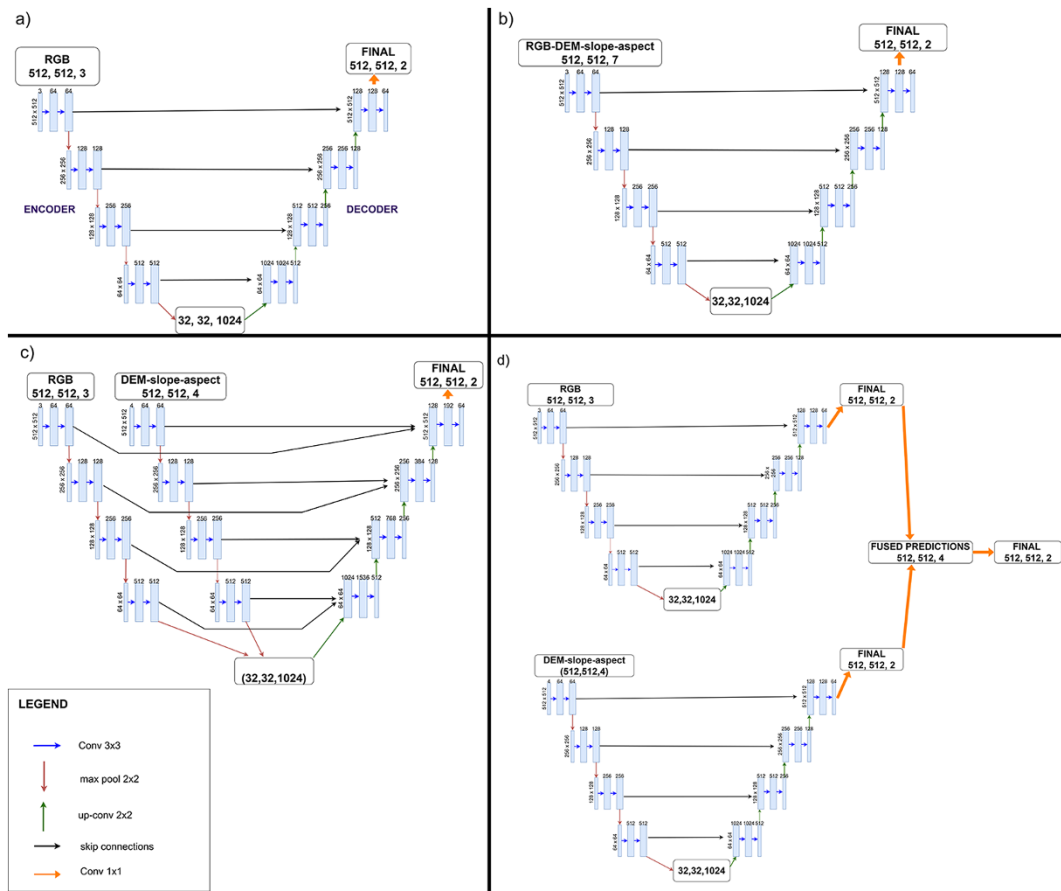


Figure 3. U-Net-based fusion architectures for terrace detection: (a) RGB-only baseline; (b) early fusion (input-level integration); (c) intermediate fusion (feature-level integration); (d) late fusion (decision-level integration).

(Rezatofghi et al. 2019). We used the Adam optimizer (initial learning rate 1×10^{-3}), a batch size of eight, and trained for up to 150 epochs with early stopping based on validation IoU. Each configuration was run 10 times with different random initializations to check that the results were stable. All experiments were carried out in Google Colab Pro+ on an NVIDIA A100 GPU (40 GB VRAM).

Model evaluation followed a two-stage design to ensure transparent methodology and independent validation. The author manually digitized terraces within 16 predefined areas, producing 256 high-resolution image patches (512×512 px, 30 cm). Of these, 10% (26 patches) were reserved as unseen test data. Model predictions in these test areas were compared against the author's manual digitization, which served as the ground-truth reference.

For additional archaeological context, model outputs were also compared with Demirciler (2014), who produced an independent terrace inventory through conventional field survey and visual interpretation. This dataset was not used for model training or internal metric calculation; rather, the same evaluation measures (accuracy, precision, recall, IoU, F1) were applied post hoc to quantify the degree of spatial agreement between automated detection and established expert mapping.

The purpose of this comparison was not to assess “performance” against another human mapper but to examine how closely automated detections align with existing archaeological knowledge. This step provides a measure of interpretive consistency and highlights areas of divergence that warrant further field investigation.

Table 4. Comparative Performance Summary of All Approaches.

Approach	Train Acc	Train IoU	Val Acc	Val IoU	Test Acc	Test IoU	Strategy
RGB-only baseline	0.849	0.737	0.856	0.749	0.843	0.729	Spectral-only
Early fusion	0.875	0.776	0.884	0.810	0.859	0.754	Input-level integration
Intermediate fusion	0.865	0.759	0.878	0.796	0.852	0.745	Feature-level integration
Late fusion	0.855	0.690	0.855	0.696	0.846	0.678	Decision-level integration

All analyses involving Demirciler's dataset were conducted only within the unseen test zones, ensuring complete independence between training and comparison data. Demirciler's work was used solely as a published archaeological reference, with no overlap in data production or authorship.

Segmentation performance was evaluated using IoU (Equation 4), accuracy (Equation 5), precision (Equation 6), recall (Equation 7), and F1-score (Equation 8). IoU (Jaccard Index) quantifies the spatial overlap between predicted terrace pixels and ground-truth labels, representing the primary segmentation metric. Accuracy reflects overall correctness, while precision, recall, and F1-score provide complementary measures of false-positive and false-negative tendencies.

$$\text{IoU} = \text{TP} / (\text{TP} + \text{FP} + \text{FN}) \quad (4)$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (5)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (6)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (7)$$

$$\text{F1 - score} = 2 \times \text{TP} / (2 \times \text{TP} + \text{FP} + \text{FN}) \quad (8)$$

where TP, FP, FN, and TN represent true positive, false positive, false negative, and true negative predictions, respectively.

Results

Fusion Strategy Comparison

Comparative evaluation of four U-Net-based architectures using standard segmentation metrics (Table 4) shows that early fusion achieved optimal performance (test IoU: 0.754), followed by intermediate fusion (0.745), RGB-only baseline (0.729), and late fusion (0.678). The RGB-only baseline's strong performance indicates that spectral information alone provides substantial terrace detection capability. However, topographic data integration at input level (early fusion) or feature level (intermediate fusion) yielded measurable improvements over spectral-only detection.

Early fusion delivered the highest performance (IoU = 0.754, Accuracy = 0.859), showing that combining spectral and topographic information at the input stage produced the most reliable terrace segmentation (Table 4). The seven-channel input, merging normalized RGB and terrain derivatives, improved detection stability and accuracy. Intermediate fusion yielded similar but slightly lower scores, suggesting that processing both data types together from the start was more effective than merging them in deeper layers. The RGB-only model also performed well (IoU = 0.729), confirming that imagery alone carries strong discriminative power for terrace mapping. Yet, the inclusion of elevation and slope data still provided measurable gains, particularly in complex terrain. Late fusion performed worst (IoU = 0.678) and showed greater variability, likely because integrating information only at the decision stage limited the model's ability to learn complementary patterns.

Given its balanced and consistent results across all metrics, the early fusion model was selected for detailed analysis as the most effective configuration (Table 5). The early fusion model achieved the highest test performance across 10 runs (IoU = 0.754, Accuracy = 0.859; Table 5). Adding topographic inputs

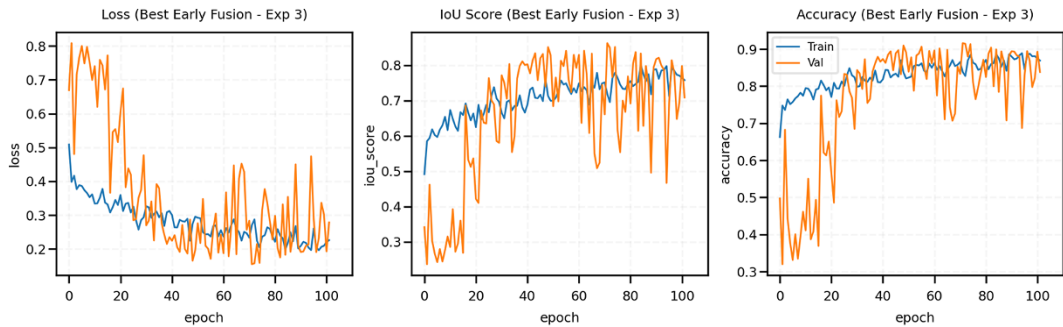


Figure 4. Training and validation curves for early fusion (best run).

improved results over the RGB-only baseline (IoU = 0.729, Accuracy = 0.843), showing that elevation and slope contribute meaningfully to terrace detection.

Among the Monte Carlo runs, Experiment 3 yielded the best validation score (IoU = 0.862) and was used for detailed curve analysis (Figure 4). Training stabilized after the first 10 epochs, and validation IoU occasionally surpassed training values, indicating good generalization and effective learning of spectral–topographic relations.

Across different initializations, test IoU values ranged from 0.647 to 0.845, confirming consistent behavior. For archaeological terrace detection, the early fusion approach offered a strong balance between accuracy, stability, and computational efficiency—surpassing spectral-only models while remaining lightweight.

Test Predictions

Qualitative evaluation further supported these results. Figure 5 presents representative examples from the 26 test patches (10% of the dataset), showing accurate delineation of terrace boundaries under various terrain and preservation conditions. Predicted masks closely matched ground-truth edges in well-preserved areas (Figure 5a–b) and successfully captured terrace traces in more complex, vegetated zones (Figure 5c–e). Combining spectral and topographic information improved reliability across diverse landscapes, demonstrating the potential of deep learning for large-scale terrace mapping.

Automated Terrace Mapping at Peninsula Scale

Following test validation, the early fusion model was applied to the full 193 km² Bozburun Peninsula to evaluate scalability for landscape-wide mapping. The area was divided into 40,304 overlapping tiles (512 × 512 px, 25% overlap) to prevent edge artifacts and ensure seamless predictions. The same preprocessing pipeline used in training was adopted, combining normalized RGB imagery with DEM-derived topographic layers (elevation, slope, aspect_sin, aspect_cos).

The large-scale mapping identified 2,517 ha of ancient agricultural terraces—about 13% of the peninsula. Figure 6 shows their spatial distribution, revealing broad coverage across diverse topographic zones. This represents the first comprehensive documentation of Hellenistic period terrace systems across the peninsula, providing landscape-scale completeness that would be impossible through traditional survey methods.

Demirciler (2014) mapped terrace zones using manual photointerpretation and GIS delineation, resulting in generalized terrace areas based on visual interpretation. In contrast, the present model applies pixel-level semantic segmentation to detect discrete terrace remnants. Therefore, differences between the two datasets reflect methodological resolution rather than disagreement in classification.

To examine detection differences locally, a qualitative visual comparison was carried out using representative subsets across the peninsula. Three main outcomes were observed: (a–b) terraces detected only by the AI mo (c–d) terraces recorded only in archaeological mapping and (e–f) terraces identified by both methods (Figure 7). These examples were drawn from areas outside the test dataset to maintain

Table 5. Early Fusion Performance Metrics (10 Runs).

	Training IoU	Validation IoU	Training Accuracy	Validation Accuracy	Test Loss	Test IoU	TestAccuracy
1	0.739	0.804	0.854	0.863	0.155	0.845	0.916
2	0.784	0.762	0.881	0.881	0.231	0.769	0.869
3	0.718	0.862	0.838	0.916	0.248	0.752	0.858
4	0.753	0.795	0.864	0.867	0.219	0.781	0.877
5	0.815	0.820	0.902	0.899	0.313	0.687	0.815
6	0.769	0.769	0.868	0.861	0.353	0.647	0.786
7	0.839	0.807	0.917	0.859	0.220	0.780	0.877
8	0.773	0.812	0.869	0.887	0.252	0.748	0.856
9	0.789	0.823	0.879	0.896	0.197	0.803	0.891
10	0.778	0.842	0.878	0.908	0.272	0.728	0.843
Avg.	0.776	0.810	0.875	0.884	0.246	0.754	0.859

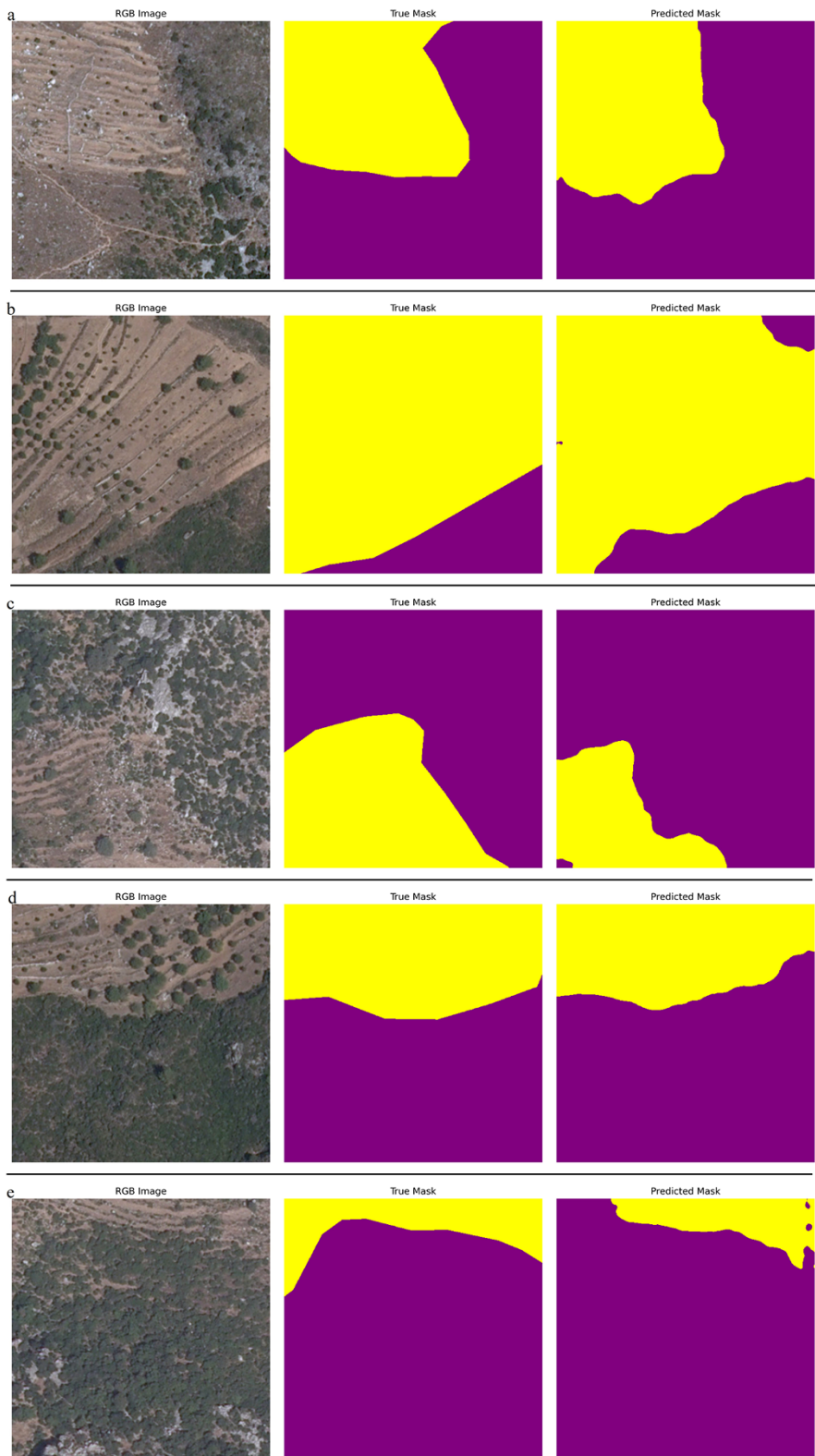


Figure 5. Test predictions from early fusion model. Representative examples (a–e) showing (left to right): RGB input; ground-truth mask; predicted mask. Yellow indicates terraced areas; purple represents background.



Figure 6. Full-extent terrace mapping results: (a) terraces predicted by the early fusion AI model and (b) terraces digitized by Demirciler (2014). Both panels share identical symbology and scale to enable visual comparison of terrace coverage and distribution patterns. Although both datasets cover largely overlapping regions, their mapping boundaries are not identical; the extent of their intersection used for quantitative comparison is shown in Figure 8.

methodological independence and to illustrate how the two mapping approaches align or diverge across different landscapes and preservation contexts.

Spatial Analysis of Detected Terraces

To examine how detected terraces align with archaeological expectations, results were compared within the geographic overlap between the AI model's coverage and Demirciler's (2014) mapping extent. This shared area enabled direct comparison between automated and expert mapping. Within this zone, elevation, slope, and aspect distributions were analyzed to identify spatial agreement and divergence, treating Demirciler's dataset as an independent reference rather than a performance benchmark (Figure 8; Tables 6–8).

The comparison focused on spatial relationships rather than accuracy testing. Elevation distributions showed strong consistency: 87.33% of terraces in Demirciler's mapping and 89.77% in the model's

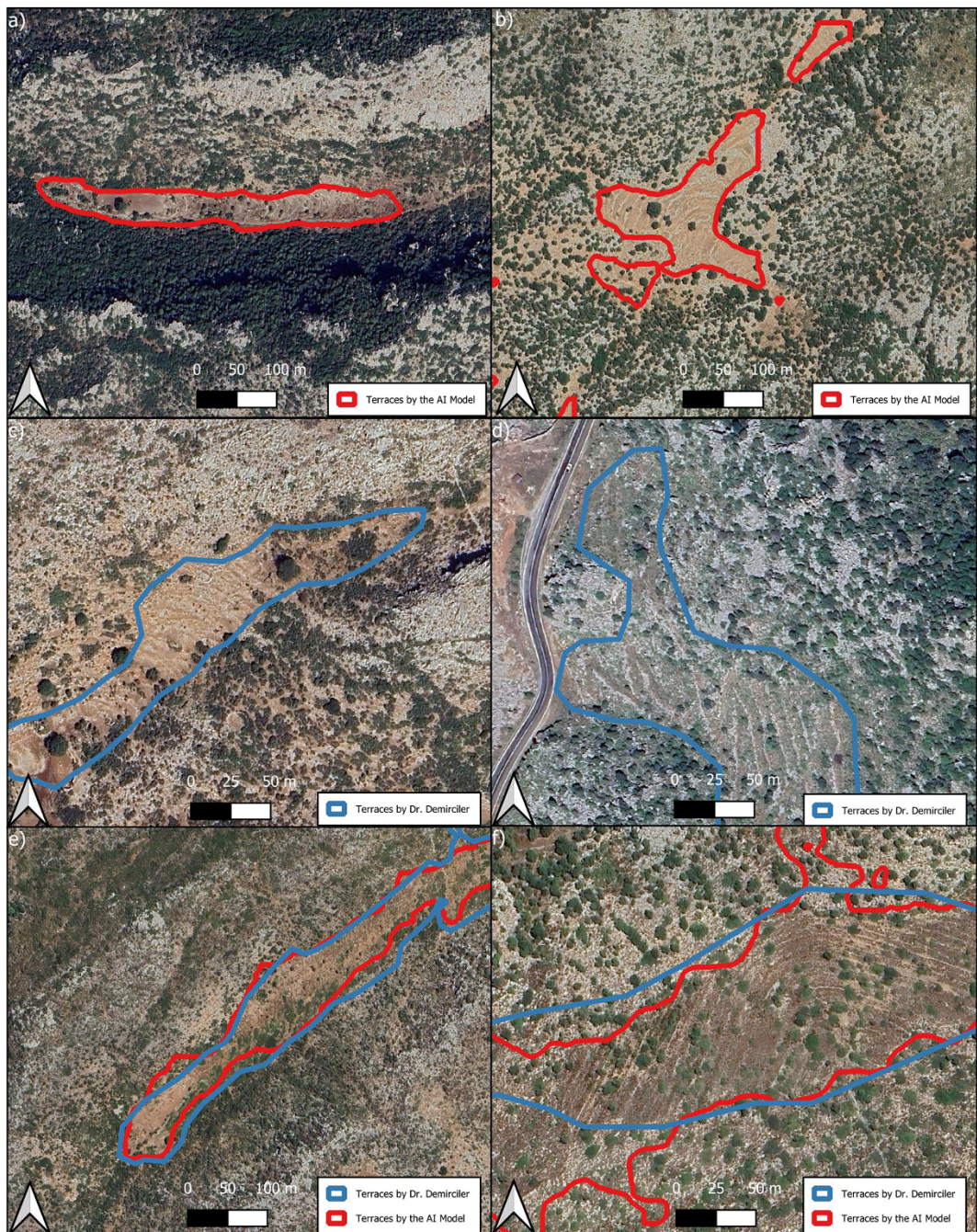


Figure 7. Representative examples showing local detection outcomes across the Bozburun Peninsula: (a–b) terraces detected only by the AI model (red); (c–d) terraces identified only by archaeological mapping (blue); (e–f) terraces detected by both methods (red and blue overlapping). All examples are located outside the test areas to maintain independence from the quantitative evaluation.

prediction occur below 300 m. Both datasets identified few terraces above 500 m (0.58% vs. 0.06%), confirming the upper limits of ancient terrace agriculture in the region (Table 6).

Slope distributions revealed complementary detection strengths between the two approaches. Both identified moderate slopes (10° – 30°) as the dominant terrace zones, confirming expected construction

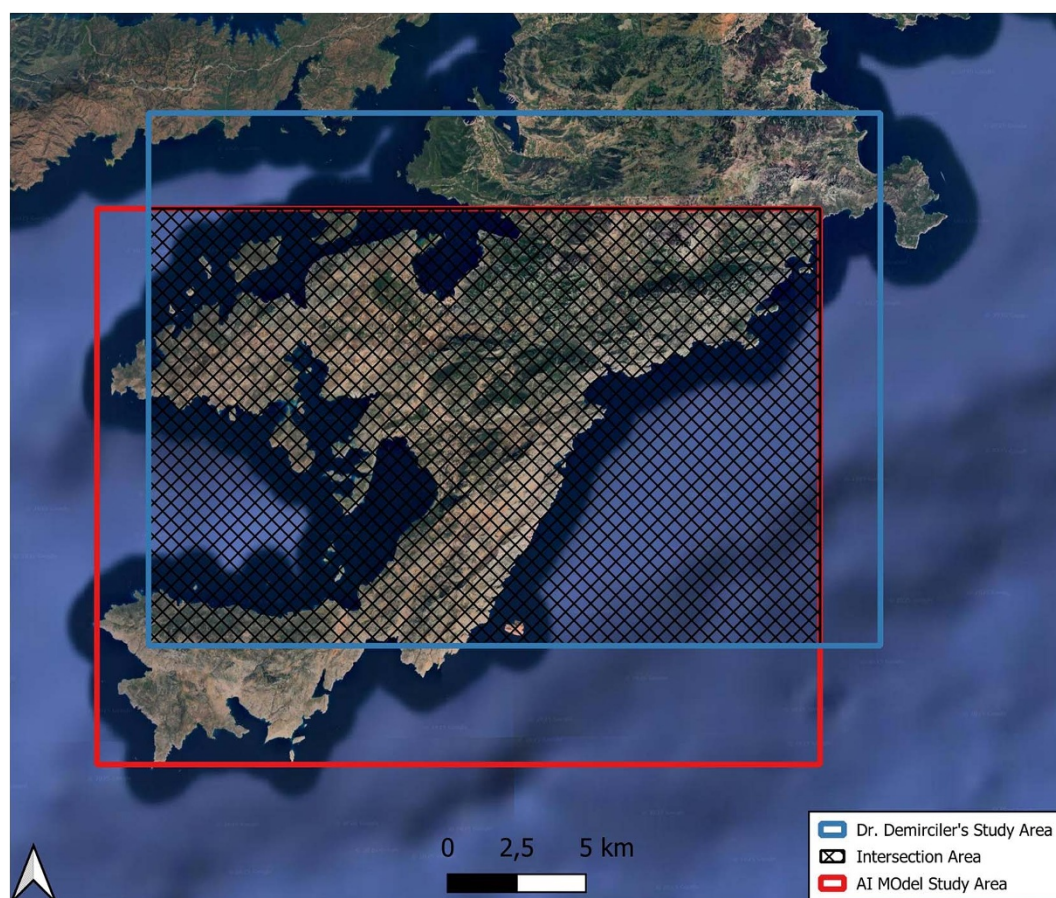


Figure 8. Geographic overlap between the model's full-extent prediction and Demirciler's (2014) terrace mapping; the hatched intersection indicates the area used for spatial statistics (elevation, slope, aspect).

preferences of ancient systems (Table 7). The model showed slightly higher sensitivity on gentler gradients ($<10^\circ$: 21.42% vs. 18.39%), likely detecting faint or eroded terraces that are less visible in traditional mapping. Archaeological mapping recorded slightly more terraces on steep slopes ($>30^\circ$: 7.66% vs. 7.46%), possibly reflecting differences in visual interpretation at higher relief. The small share of terraces detected above 500 m likely results from modern terrace construction, natural benches, or limited field verification rather than model error.

Aspect distributions showed similar general patterns with minor directional differences (Table 8). The model detected more terraces on north-facing slopes (18.81% vs. 15.55%) and fewer on southwest- and west-facing slopes, patterns that likely reflect differences in illumination, preservation, and vegetation cover.

These variations carry archaeological significance. Terraces detected only by the AI model—mainly at lower elevations and gentler slopes—may correspond to eroded or vegetation-covered structures that are difficult to see in aerial imagery. In contrast, terraces mapped only by Demirciler often occur on steeper or irregular benches where automated detection struggled to maintain boundary continuity. Slopes over 30° were retained in the analysis, as they may still represent degraded terrace remnants and offer an insight into past landscape stability. Although some terraces may coincide with modern agricultural reuse, distinguishing ancient from contemporary structures lies beyond this study's scope and requires field verification. The strong correspondence in aspect distributions supports interpretations that ancient terraces were aligned to optimize sunlight and wind protection. Overall, the comparison shows that

Table 6. Comparison of Elevation Distributions.

Elevation Intervals (m)	Percentage (%) of Demirciler	Percentage (%) of the Model
<300m	87.33	89.77
300–500m	12.09	10.17
500–800m	0.58	0.06
>800m	0.00	0.00

Table 7. Comparison of Slope Distributions.

Slope Intervals (degrees)	Percentage (%) of Demirciler	Percentage (%) of the Model
<10°	18.39	21.42
10–20°	44.30	40.98
20–30°	29.65	30.14
>30°	7.66	7.46

Table 8. Comparison of Aspects.

Aspect (Direction)	Percentage (%) of Demirciler	Percentage (%) of the Model
North	15.55	18.81
Northeast	8.71	11.39
East	10.13	12.48
Southeast	12.24	13.60
South	14.44	13.00
Southwest	11.03	7.37
West	12.70	9.31
Northwest	15.21	14.04

Table 9. Performance Comparison between AI Model and Expert Documentation.

	Terraces predicted by the model	Terraces digitized by Demirciler (2014)
Accuracy	0.843	0.862
Precision	0.874	0.793
Recall	0.766	0.943
IOU	0.690	0.757
F1	0.816	0.862

automated terrace detection reproduces meaningful archaeological patterns while highlighting areas that merit future ground validation.

Model Performance Assessment

For comparison with archaeological documentation (Demirciler 2014), performance metrics were calculated using the 23 test patches located within the spatial overlap zone, excluding three southern patches outside this area (Figure 8).

The comparison highlights complementary strengths rather than a performance hierarchy between approaches. The early fusion model achieved higher precision, meaning it produced fewer false positives, while archaeological mapping showed higher recall, reflecting expert recognition of subtle or eroded terraces (Table 9). Together, these results suggest that automated and expert mapping capture different but equally valuable aspects of terrace distribution.

Differences in the metrics reflect the complementary value of each approach. The model's higher precision indicates reliable terrace detection where features are clearly visible in the imagery, while higher recall in archaeological documentation underscores the continuing importance of expert interpretation for recognizing degraded or vegetation-obscured remains. Rather than suggesting a hierarchy of performance, these findings demonstrate that automated mapping can enhance—but not replace—expert knowledge, especially in complex or eroded terrain.

Conclusions

This study presents a deep learning-based approach for automated detection of ancient agricultural terraces in Mediterranean archaeological landscapes. Among four U-Net-based architectures tested, early fusion—integrating spectral and topographic data at the input level—achieved the best results, with an IoU of 75.4% and an accuracy of 85.9% across the 193 km² Bozburun Peninsula. Early fusion outperformed intermediate fusion (74.5% IoU), late fusion (67.8%), and the RGB-only baseline (72.9%), confirming that input-level integration supports more effective interaction between spectral and topographic information. The preprocessing steps, including circular encoding of aspect values and normalization by global maxima, contributed to stable model behavior. Monte Carlo evaluation with 10 random initializations provided consistent results and established replicable benchmarks for archaeological feature detection in Mediterranean environments.

The identification of 2,517 ha of terraces represents extensive documentation of agricultural systems spanning multiple historical phases. Spatial analysis revealed distribution patterns consistent with ancient site-selection strategies: 89.77% of terraces occur below 300 m elevation, mostly on 10°–30° slopes (71.14%), with a preference for north-facing orientations (18.81%). These findings indicate deliberate adaptation to terrain conditions that optimized soil stability, accessibility, and agricultural productivity. Comparison with archaeological mapping showed strong spatial agreement, with both approaches identifying similar distribution trends. The AI model achieved higher precision (87.4% vs. 79.3%), while expert mapping achieved higher recall (94.3% vs. 76.6%), confirming their complementary nature.

Deep learning thus serves as an effective complement to traditional survey methods, offering scalable tools for regional documentation. Processing the full 193 km² study area demonstrates its potential for heritage management, particularly where fieldwork is limited by time or resources. Nonetheless, limitations remain, including reliance on high-quality training data and sensitivity to vegetation and erosion. Future work should explore lighter network architectures, transfer learning for cross-regional generalization, and multitemporal analysis to monitor terrace degradation. Integrating terrace detection with other archaeological features may further improve interpretations of land use and settlement dynamics, while hybrid workflows combining automated detection with expert validation appear most effective for balancing efficiency and accuracy.

The mapped terrace distributions provide a quantitative basis for conservation planning and long-term monitoring. Combining computational and archaeological perspectives offers a more systematic understanding of ancient agricultural practices and supports sustainable management of Mediterranean cultural landscapes. Terrace concentrations correspond closely with known Hellenistic rural production zones, such as those near Tymnos, suggesting continuity between agricultural infrastructure and socioeconomic organization. These patterns align with areas historically associated with wine and olive oil production under Rhodian administration, offering quantifiable evidence of agricultural intensification in southwestern Anatolia.

When both accuracy and recall are considered, early fusion provides only modest improvement relative to RGB-only models—a practical outcome for large-area mapping where additional data sources may not always justify their cost.

Key Takeaways for Archaeological Practice

- RGB-only models achieve near-optimal terrace detection performance in this Mediterranean context, reducing reliance on additional topographic data sources.

- When available, DEM-derived features are most beneficial when integrated via early fusion; intermediate and late fusion provide limited advantage.
- Full-coverage implementation across 193 km² demonstrates the feasibility of deep learning for large-area cultural heritage documentation.

Acknowledgments. This article is based on research originally conducted as part of the author's doctoral dissertation at the Graduate School of Social Sciences Settlement Archaeology program at Middle East Technical University. The author thanks Prof. Dr. Burcu Erciyas for supervision, Assist. Prof. Dr. Volkan Demirciler, and Assoc. Prof. Dr. E. Deniz Oğuz-Kırca for valuable input. Data access was provided by the General Directorate of Mapping, Turkey (2024). The author acknowledges the use of OpenAI's ChatGPT (GPT-5) for language editing and phrasing assistance during manuscript preparation. The tool was not used for data analysis, figure generation, or interpretation.

Funding Statement. This research received no specific grant from any funding agency or organization.

Data Availability Statement. RGB orthophotos and DEMs were obtained from the General Directorate of Mapping, Turkey, under research authorization. The complete source code is available at <https://doi.org/10.5281/zenodo.16905405>. Trained model weights are available on request. Demirciler's (2014) dataset was accessed with the author's permission for academic comparison only and used solely as an independent archaeological reference for spatial agreement analysis.

Competing Interests. The author declares none.

References Cited

- Agapiou, Athos, Dimitrios D. Alexakis, and Diofantos G. Hadjimitsis. 2019. Potential of Virtual Earth Observation Constellations in Archaeological Research. *Sensors* 19(19):4066.
- Boulahia, Said Yacine, Abdellah Amamra, Mohamed Reda Madi, and Samir Daikh. 2021. Early, Intermediate, and Late Fusion Strategies for Robust Deep Learning-Based Multimodal Action Recognition. *Machine Vision and Applications* 32(6):121–132.
- Capolupo, Alessandra, Lammert Kooistra, and Lorenzo Boccia. 2018. A Novel Approach for Detecting Agricultural Terraced Landscapes from Historical and Contemporaneous Photogrammetric Aerial Photos. *International Journal of Applied Earth Observation and Geoinformation* 73:800–810.
- Cigliç, Rok, Anže Glušič, Luka Štanut, and Luka Črt Zajc. 2024. Towards the Deep Learning Recognition of Cultivated Terraces Based on LiDAR Data: The Case of Slovenia. *Moravian Geographical Reports* 32(1):66–78.
- Cucchiari, Stefano, Daniele J. Fallu, Haitao Zhang, Kevin Walsh, Kristof Van Oost, Antony G. Brown, and Paolo Tarolli. 2020. Multiplatform-SfM and TLS Data Fusion for Monitoring Agricultural Terraces in Complex Topographic and Landcover Conditions. *Remote Sensing* 12(12):1946.
- Demirciler, Volkan. 2014. Agricultural Terraces and Farmsteads of Bozburun Peninsula in Antiquity. PhD dissertation, Graduate School of Social Sciences, Middle East Technical University, Ankara, Turkey.
- Demirciler, Volkan. 2017. Photogrammetrical Applications and GIS Analyses of Ancient Agricultural Terraces in Bozburun Peninsula. In *Proceedings of the International Symposium on GIS Applications in Geography & Geosciences*, edited by Korkmaz Atalay, pp. 21–32. Çanakkale Onsekiz Mart University, Çanakkale, Turkey.
- Demirciler, Volkan, E. Deniz Oğuz-Kırca, Özer Sarı, and Cengiz Bakan. 2023. Settlement Systems and Agricultural Organizations from the Early Iron Age to Late Antiquity in the Bozburun Peninsula Archaeological Surface Survey 2022 Studies. In *43. Kazı Sonuçları Toplantısı 2. Cilt*, pp. 483–502. Kültür Varlıkları ve Müzeler Genel Müdürlüğü Yayınları, Ankara, Turkey.
- Demirciler, Volkan, E. Deniz Oğuz-Kırca, Gamze V. Toprak, Zeynep Karahan, and Çiğdem Kılıç. 2022. Archaeological Survey of Settlement Systems and Agricultural Organizations from the Early Iron Age to Late Antiquity in Bozburun Peninsula: 2021 Campaign [in Turkish]. In *38. Araştırma Sonuçları Toplantısı 2. Cilt*, pp. 117–138. Kültür Varlıkları ve Müzeler Genel Müdürlüğü Yayınları, Ankara, Turkey.
- Do, Huy Tho, Vijay Raghavan, and Go Yonezawa. 2019. Pixel-Based and Object-Based Terrace Extraction Using Feed-Forward Deep Neural Network. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-3/W1:1–7.
- Figueiredo, Nelson, António Neto, António Cunha, João Sousa, and António Sousa. 2022. Deep Learning Approach for Terrace Vineyards Detection from Google Earth Satellite Imagery. In *IGARSS 2022 – 2022 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5824–5827. IEEE, Piscataway, New Jersey.
- Figueiredo, Nelson, Luís Pádua, António Cunha, João J. Sousa, and António Sousa. 2023. Exploratory Approach for Automatic Detection of Vine Rows in Terrace Vineyards. *Procedia Computer Science* 217:139–144.
- Gravel-Miguel, Claudine, Grant Snitker, Jayde N. Hirniak, Katherine Peck, and Alex Fetterhoff. 2025. Semantic Segmentation of Archaeological Features on Public Lands: Case Study of Historical Cotton Terraces within the Piedmont National Wildlife Refuge, Georgia, USA. *Advances in Archaeological Practice* 13(2):291–314.
- Guttmann-Bond, Emma. 2019. *Reinventing Sustainability: How Archaeology Can Save the Planet*. Oxbow Books, Oxford.
- Krahtopoulou, Anastasia, and Charles Frederick. 2008. The Stratigraphic Implications of Long-Term Terrace Agriculture in Dynamic Landscapes: Polycyclic Terracing from Kythera Island, Greece. *Geoarchaeology* 23(4):550–585.

- Kuban, Zeynep. 2008. Agricultural Units Identified during the Surface Survey of the Sanctuary of Kiran Lake in the Carian Chersonese (Bozburun Peninsula). In *Olive Oil and Wine Production in Anatolia during Antiquity*, edited by Ümit Aydınoğlu and Ahmet K. Şenol, pp. 213–225. Research Center of Cilician Archaeology, Mersin, Turkey.
- Lu, Yalin, Xue Li, Liang Xin, Hongqi Song, and Xia Wang. 2023. Mapping the Terraces on the Loess Plateau Based on a Deep Learning Model at 1.89 m Resolution. *Scientific Data* 10:115.
- Qiu, Kevin, Lina E. Budde, Denis Bulatov, and Dorota Iwaszczuk. 2022. Exploring Fusion Techniques in U-Net and DeepLabV3 Architectures for Multi-Modal Land Cover Classification. *Proceedings of SPIE* 12268:190–200.
- Raju, Chaitanya Sagar, Bharath Chandra Neelapu, Rahmatullah Hussain Laskar, and Ghulam Muhammad. 2025. Analysis of Multimodal Fusion Strategies in Deep Learning for Ischemic Stroke Lesion Segmentation on Computed Tomography Perfusion Data. *Multimedia Tools and Applications* 84(10):7493–7518.
- Rezatofghi, Hamid, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. 2019. Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 658–666. IEEE, Piscataway, New Jersey.
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, edited by Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, pp. 234–241. Lecture Notes in Computer Science Vol. 9351. Springer, Cham, Switzerland.
- Santos, Luís, Francisco N. Santos, Vitor Filipe, and Pramod Shinde. 2019. Vineyard Segmentation from Satellite Imagery. In *Progress in Artificial Intelligence*, edited by Paulo Moura Oliveira, Paulo Novais, and Luís Paulo Reis, pp. 109–120. Springer, Cham, Switzerland.
- Sofia, Giulia, Federico Marinello, and Paolo Tarolli. 2014. A New Landscape Metric for the Identification of Terraced Sites: The Slope Local Length of Auto-Correlation (SLLAC). *ISPRS Journal of Photogrammetry and Remote Sensing* 96:123–133.
- Tubog, Mark Vincent, Karim Emsellem, and Stéphane Bouissou. 2025. Detection of Agricultural Terrace Platforms Using Machine Learning. *Land* 14(5):962.
- Wang, Yimin, Chao Liu, Arnon Karnieli, and Xiao Xiang Zhu. 2022. Deep Semantic Model Fusion for Ancient Agricultural Terrace Detection. In *Proceedings of the 2022 IEEE International Conference on Big Data*, pp. 4888–4892. IEEE, Piscataway, New Jersey.
- Xu, Qingsong S., and Yi-Zeng Liang. 2001. Monte Carlo Cross Validation. *Chemometrics and Intelligent Laboratory Systems* 56(1):1–11.
- Zhao, Fan, Liang Xiong, Cheng Wang, Hong-Rui Wang, Hui Wei, and Guoyang Tang. 2021. Terrace Mapping Using Deep Learning from Remote Sensing Images and Digital Elevation Models. *Transactions in GIS* 25(5):2438–2454.
- Zhao, Yuxi, Jiawei Zou, Shuyang Liu, and Yuan Xie. 2024. Terrace Extraction Method Based on Remote Sensing and a Novel Deep Learning Framework. *Remote Sensing* 16(9):1649.